Visualização Multi-dimensional:

Mineração Visual de Dados multidimensionais e aplicações Parte I

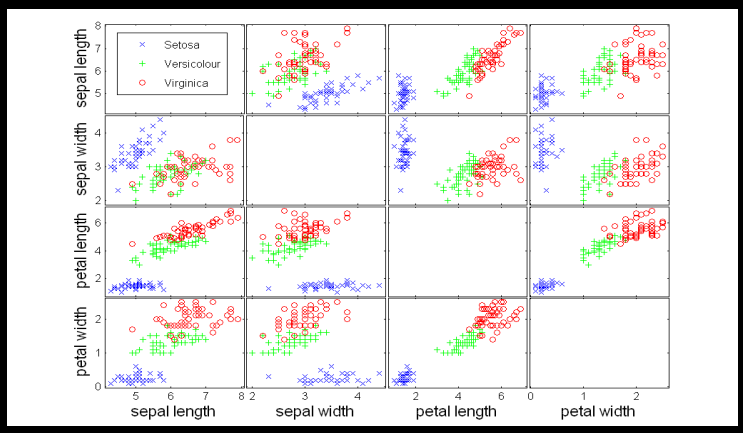
Rosane Minghim

2008-2012

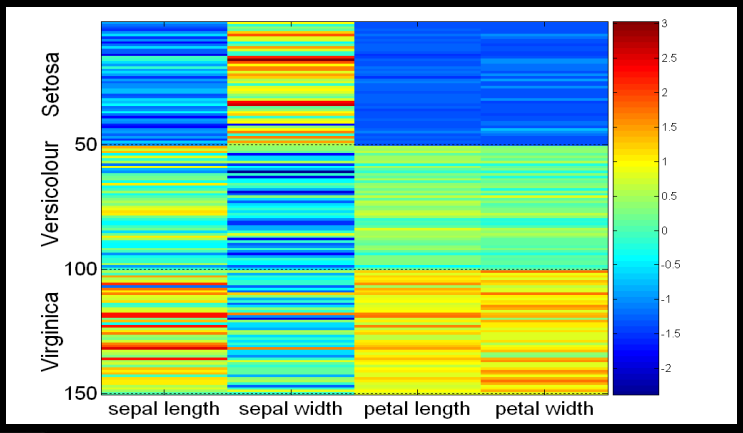
Visualização Multidimensional
Projeções
Coleções de Documentos
Coleções de Imagens

- Visualização
- Análise de Dados eMineração Visual
- Projeções
- Exemplos: Textos e Imagens
- Continuidade

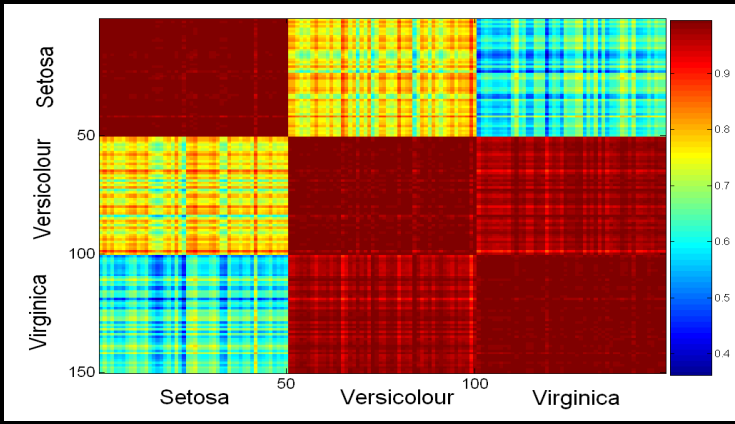
Scatter Plot Matrix



Data Matrix

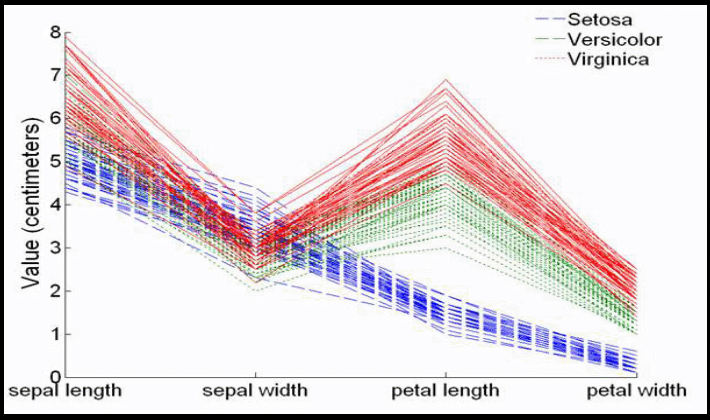


Correlation Matrix



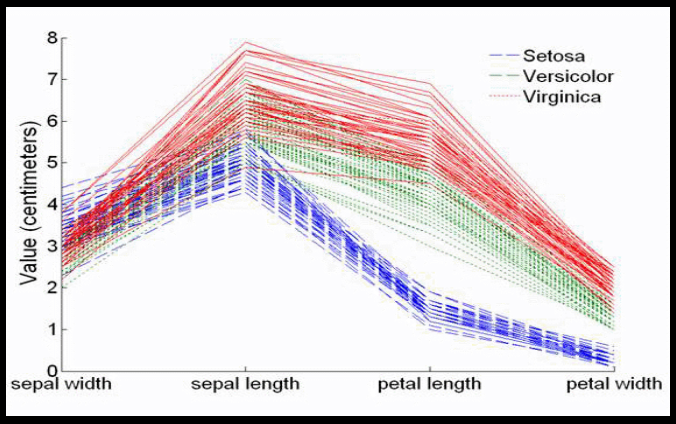
5

Parallel Coordinates



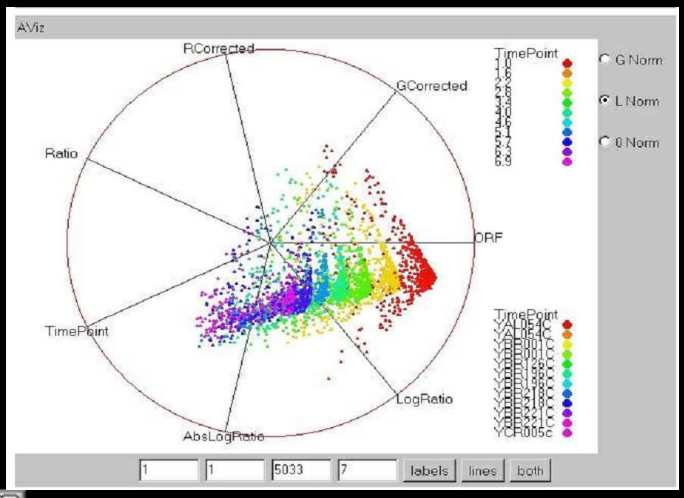
6

Parallel Coordinates



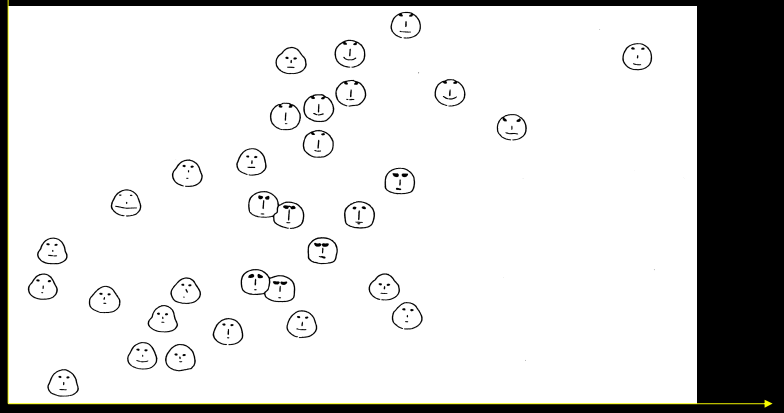
7

RadViz



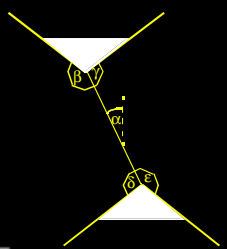
8

Chernoff Faces

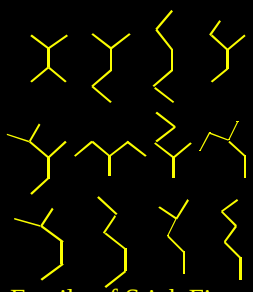


Stick Figures

- properties of a line icon
- two attributes of the data are mapped to the display axes
- the remaining attributes are mapped to the angle/length of the limbs
- texture patterns in the visualization show certain data characteristics



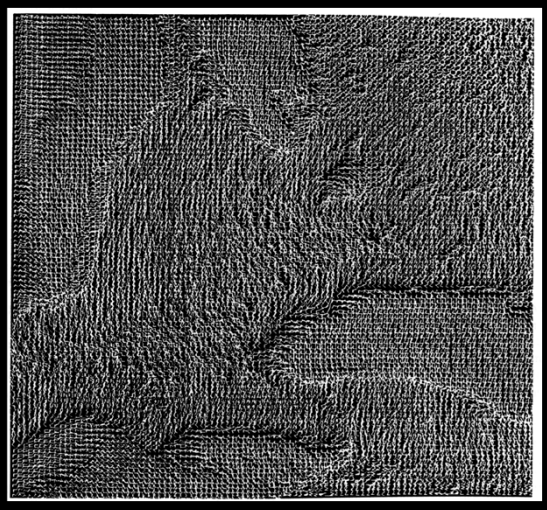
Stick Figure Icon



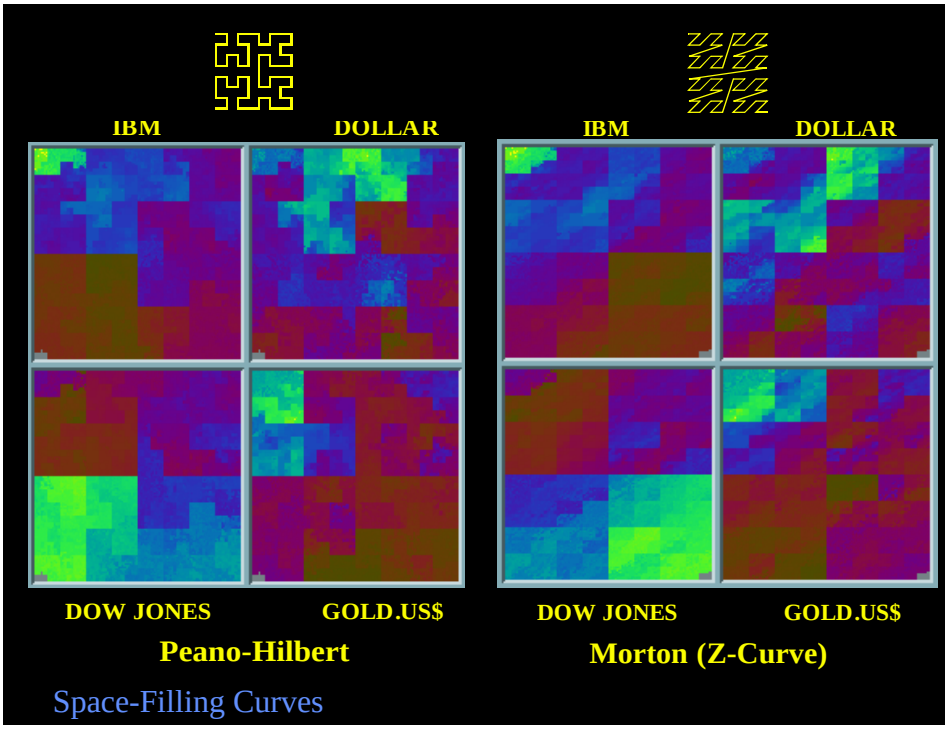
A Family of Stick Figures

Stick Figure Icon

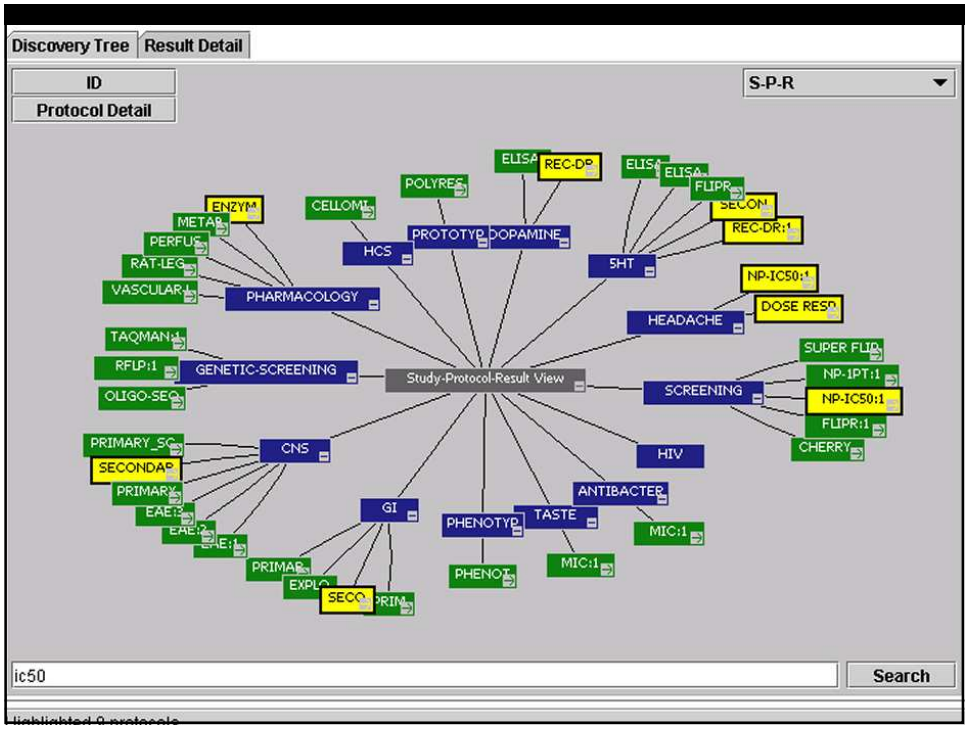
5-dimensional NOAA image data from the great lake region



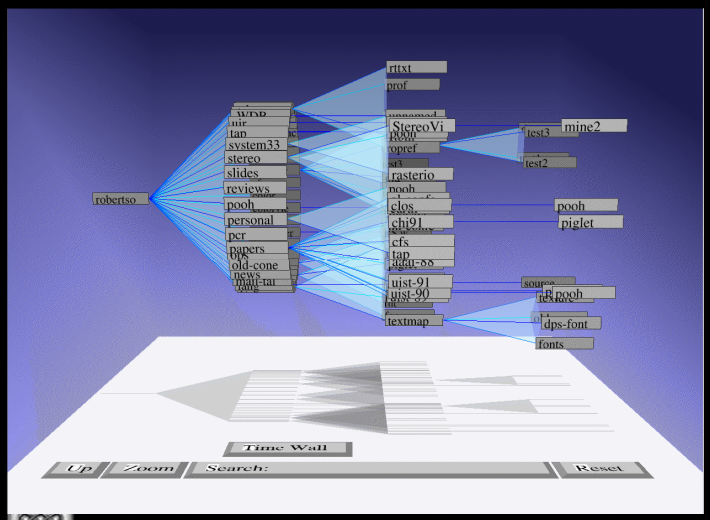
11



Space-Filling Curves



Cone Tree



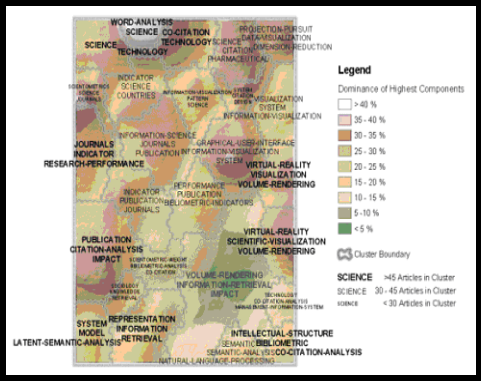
File system structure visualized as a cone tree

Animated 3D visualization of hierarchical data

14

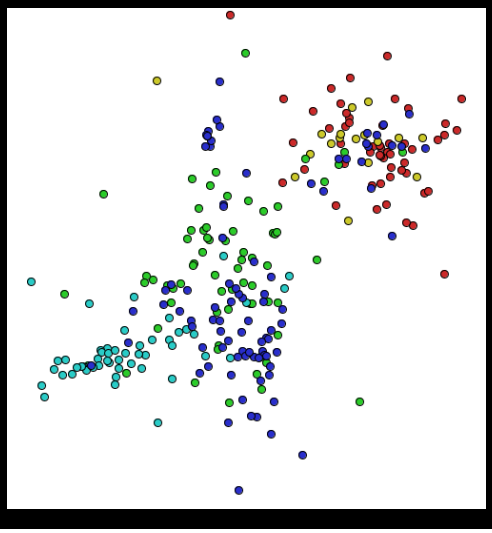
SOM based

- **Self-Organization Maps (SOMs) cartográficos (ex. Skurpin 2002)**



15

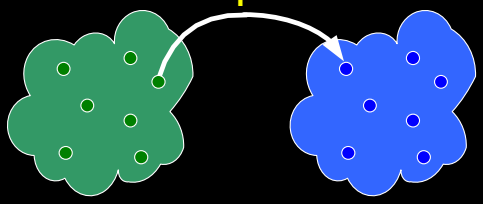
Mapeamento para o plano permitindo a exploração.
Ex: Patents surgery, drugs, molecular bio



16

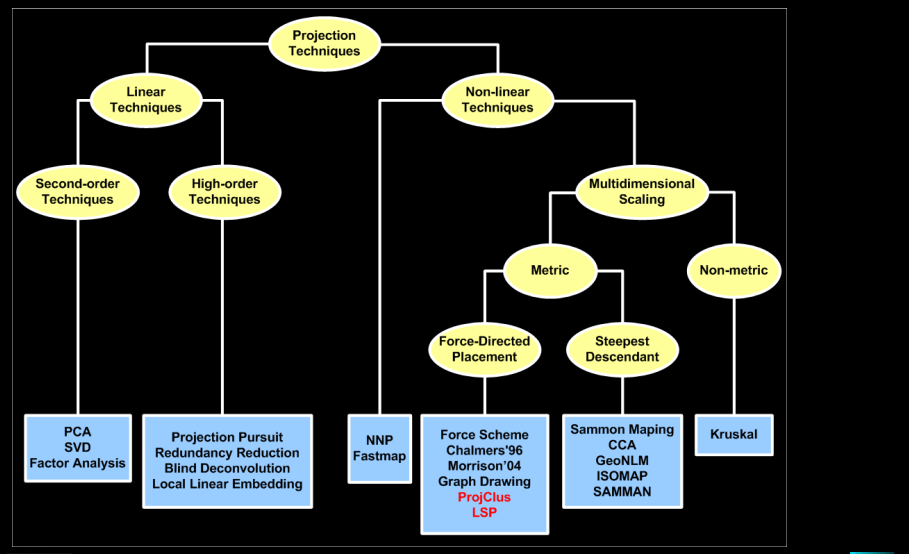
Projection Techniques

$X \in R^m$ f $Y \in R^{p=\{1,2,3\}}$



- δ: $x_i, x_j \rightarrow R$, $x_i, x_j \in X$
- d: $y_i, y_j \rightarrow R$, $y_i, y_j \in Y$
- f: $X \rightarrow Y, |\delta(x_i,x_j) - d(f(x_i), f(x_j))| \approx 0, \forall\ x_i, x_j \in X$

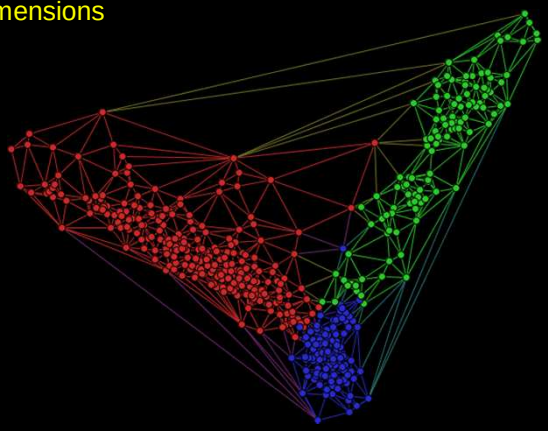
17



18

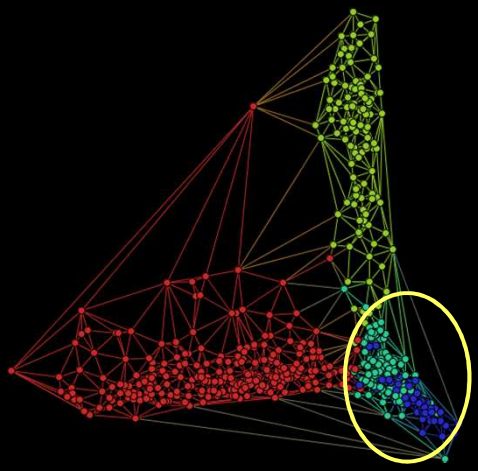
Problems PCA

390 dimensions



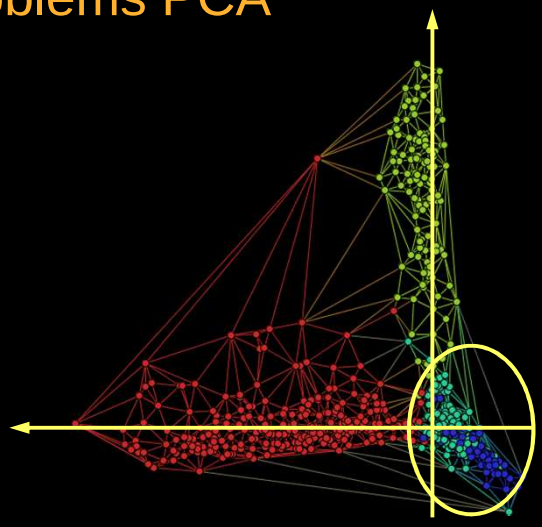
19

Problems PCA



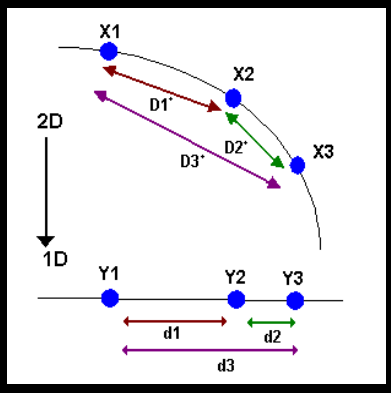
20

Problems PCA

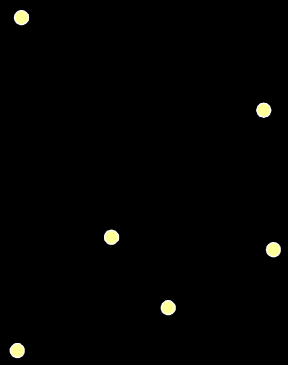


Ex: Sammon Mapping

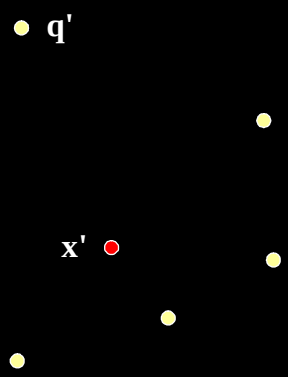
- Let **X** be the points in the original space R^n, we apply a distance measure d_{ij}* between Xi an Xj., and find **Y, the projected point**, ex. R^2 and d_{ij} the Euclidean distance between them.
- Sammon's method applies an error function to measure the target.



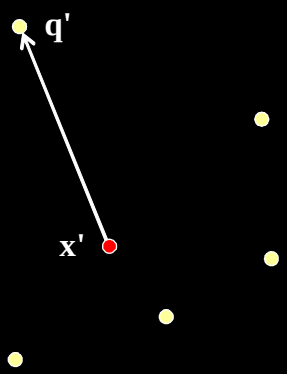
Force Based Point Placement



Force Scheme [Tejada et al., 2003]

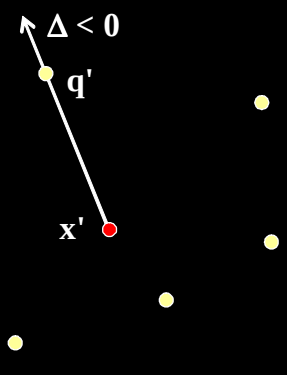


Force Scheme [Tejada et al., 2003]



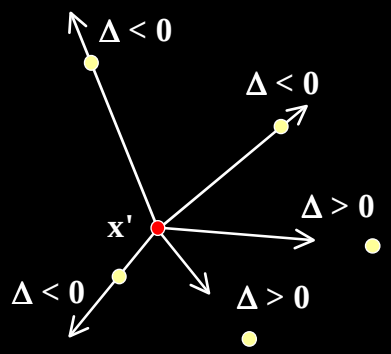
25

Force Scheme [Tejada et al., 2003]



26

Force Scheme [Tejada et al., 2003]



Force Scheme [Tejada et al., 2003]

1. Mapear cada ponto em X para um ponto no plano (fastmap, nnp, etc.)
2. Para cada ponto projetado x'
    1. Para cada ponto projetado q' ≠ x'
        1. Calcular o vetor **v** de x' para q'
        2. Mover q' na direção de **v** uma fração de Δ

$$\Delta = \frac{\delta(x,q) - \delta_{\min}}{\delta_{\max} - \delta_{\min}} - d(x',q')$$

3. Normalizar as coordenadas entre [0,1]

LSP [Paulovich et al., 2006/2008]

- Least-Square Projection (LSP)
- Idéia central: projetar um sub-conjunto de pontos e interpolar o restante
- Interpolação busca preservar a vizinhança entre os pontos
- Cada ponto é mapeado dentro do fecho convexo de seus vizinhos

29

LSP [Paulovich et al., 2006/2008]

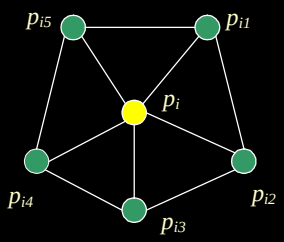
- Três passos principais
  1. Selecionar um subconjunto de pontos (*pontos de controle)* e projetar esses em R^p
  2. Determinar a vizinhança dos pontos
  3. Construir um sistema linear cujas respostas são as coordenadas cartesianas dos pontos p_i em R^p

30

LSP: Matriz Laplaciana

- Seja $V_i = \{p_{i1},\ldots,p_{iki}\}$ a vizinhança de um ponto p_i e seja c_i as coordenadas de p_i em R^p

$$c_i - \frac{1}{ki}\sum_{p_j \in V_i} c_j = 0$$



- Cada p_i será o centróide dos pontos em V_i

LSP: Matriz Laplaciana

$L\mathbf{x_1}=0,\ L\mathbf{x_2}=0,\ \ldots,\ L\mathbf{x_p}=0$

onde $x_1, x_2,\ldots, x_p$ são vetores contendo as coordenadas cartesianas dos pontos e L

- é a matriz dada por

$$Lij = \begin{cases} 1 & i = j \\ -\frac{1}{ki} & p_j \in V_i \\ 0 & \text{caso contrário} \end{cases}$$

$$\begin{pmatrix} & L & \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \ldots \\ 0 \\ 0 \end{pmatrix}$$

LSP: Adicionando os Pontos de Controle

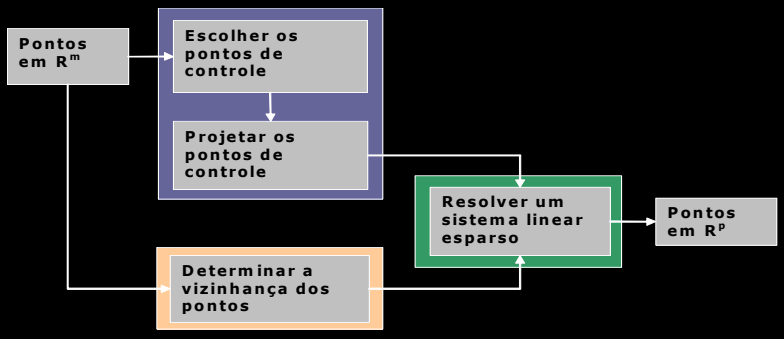
$$A = \begin{pmatrix} L \\ C \end{pmatrix} \qquad Cij = \begin{cases} 1 & p_j \text{ é um ponto de controle} \\ 0 & \text{caso contrário} \end{cases}$$

$$b_i = \begin{cases} 0 & i \le n \\ x_{p_{c_i}} & n < i \le n + nc \end{cases}$$

$$\begin{pmatrix} \mathbf{L} \\ 0\ 1\ 0\ \cdots\ 0 \\ 0\ \cdots\ 0\ 1\ 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \\ c_1 \\ c_2 \end{pmatrix}$$

33

LSP: Visão Geral



34

LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$
$v_2 = \{p_5 p_4 p_6\}$
$v_3 = \{p_1 p_5 p_6\}$
$v_4 = \{p_1 p_6\}$
$v_5 = \{p_3 p_2 p_6\}$
$v_6 = \{p_1 p_2 p_4 p_5\}$

$$L = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \end{bmatrix}$$

LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$
$v_2 = \{p_5 p_4 p_6\}$
$v_3 = \{p_1 p_5 p_6\}$
$v_4 = \{p_1 p_6\}$
$v_5 = \{p_3 p_2 p_6\}$
$v_6 = \{p_1 p_2 p_4 p_5\}$

$$L = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ & & & & & \\ & & & & & \\ & & & & & \\ & & & & & \end{bmatrix}$$

LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$
$v_2 = \{p_5 p_4 p_6\}$
$v_3 = \{p_1 p_5 p_6\}$
$v_4 = \{p_1 p_6\}$
$v_5 = \{p_3 p_2 p_6\}$
$v_6 = \{p_1 p_2 p_4 p_5\}$

$$L = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \end{bmatrix}$$

37

LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$
$v_2 = \{p_5 p_4 p_6\}$
$v_3 = \{p_1 p_5 p_6\}$
$v_4 = \{p_1 p_6\}$
$v_5 = \{p_3 p_2 p_6\}$
$v_6 = \{p_1 p_2 p_4 p_5\}$

$pc = \{p_3 p_6\}$

$$A = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

L

38

LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$
$v_2 = \{p_5 p_4 p_6\}$
$v_3 = \{p_1 p_5 p_6\}$
$v_4 = \{p_1 p_6\}$
$v_5 = \{p_3 p_2 p_6\}$
$v_6 = \{p_1 p_2 p_4 p_5\}$

$pc = \{p_3 p_6\}$

$$A = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \\ \\ \text{L} \\ \text{C} \end{matrix}$$

LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$
$v_2 = \{p_5 p_4 p_6\}$
$v_3 = \{p_1 p_5 p_6\}$
$v_4 = \{p_1 p_6\}$
$v_5 = \{p_3 p_2 p_6\}$
$v_6 = \{p_1 p_2 p_4 p_5\}$

$pc = \{p_3 p_6\}$

$$A = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ c_{x_3} \\ c_{x_6} \end{bmatrix}$$

LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$
$v_2 = \{p_5 p_4 p_6\}$
$v_3 = \{p_1 p_5 p_6\}$
$v_4 = \{p_1 p_6\}$
$v_5 = \{p_3 p_2 p_6\}$
$v_6 = \{p_1 p_2 p_4 p_5\}$

$pc = \{p_3 p_6\}$

$$A = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ c_{x_3} \\ c_{x_6} \end{bmatrix}$$

41

LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$
$v_2 = \{p_5 p_4 p_6\}$
$v_3 = \{p_1 p_5 p_6\}$
$v_4 = \{p_1 p_6\}$
$v_5 = \{p_3 p_2 p_6\}$
$v_6 = \{p_1 p_2 p_4 p_5\}$

$pc = \{p_3 p_6\}$

$$A = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ c_{y_3} \\ c_{y_6} \end{bmatrix}$$

42

LSP: Exemplo de Sistema

$v_1 = \{p_3 p_4 p_6\}$
$v_2 = \{p_5 p_4 p_6\}$
$v_3 = \{p_1 p_5 p_6\}$
$v_4 = \{p_1 p_6\}$
$v_5 = \{p_3 p_2 p_6\}$
$v_6 = \{p_1 p_2 p_4 p_5\}$

$pc = \{p_3 p_6\}$

$$A = \begin{bmatrix} 1 & 0 & -1/3 & -1/3 & 0 & -1/3 \\ 0 & 1 & 0 & -1/3 & -1/3 & -1/3 \\ -1/3 & 0 & 1 & 0 & -1/3 & -1/3 \\ -1/2 & 0 & 0 & 1 & 0 & -1/2 \\ 0 & -1/3 & -1/3 & 0 & 1 & -1/3 \\ -1/4 & -1/4 & 0 & -1/4 & -1/4 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ c_{y_3} \\ c_{y_6} \end{bmatrix}$$

LSP: Resolvendo o Sistema

- É necessário resolver $A\mathbf{x} = \mathbf{b}$
- Este sistema é resolvido usando mínimos quadrados

$$\|Ax - b\|^2$$

- A única solução analítica será

$$A^T A\mathbf{x} = A^T\mathbf{b} \quad \Rightarrow \quad \mathbf{x} = (A^T A)^{-1} A^T\mathbf{b}$$

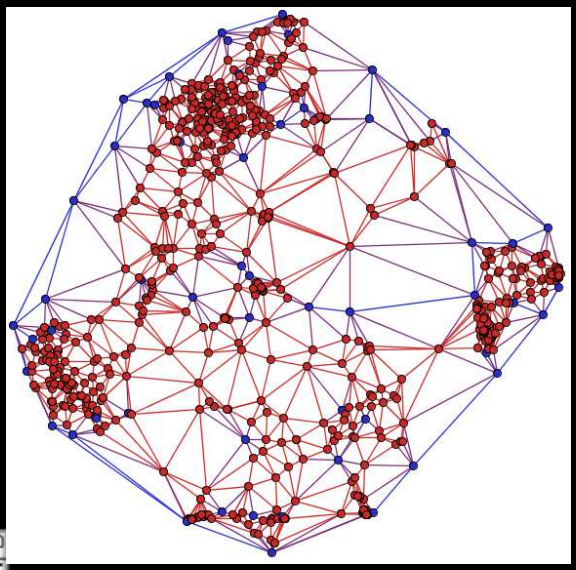
- A^TA é simétrica e esparsa e pode ser resolvida usando a fatoração de *Cholesky*

Choosing the Control Points

- In order to select the control points
  - the space R^m is split into *nc* clusters using k-medoids.
  - the control points are the medoids of each cluster

Choosing the Control Points

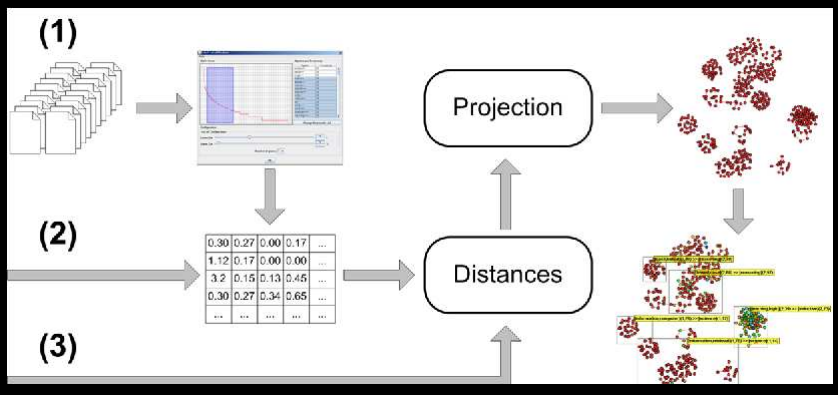
- Once the control points are chosen, these points are projected onto R^d through a fast dimensionality reduction method
  - Fast Projection (Fastmap or NNP)
  - Force Placement



Control points in blue

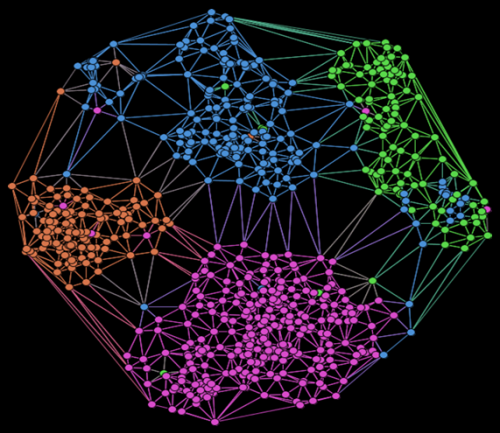
47

Content – based by Projections



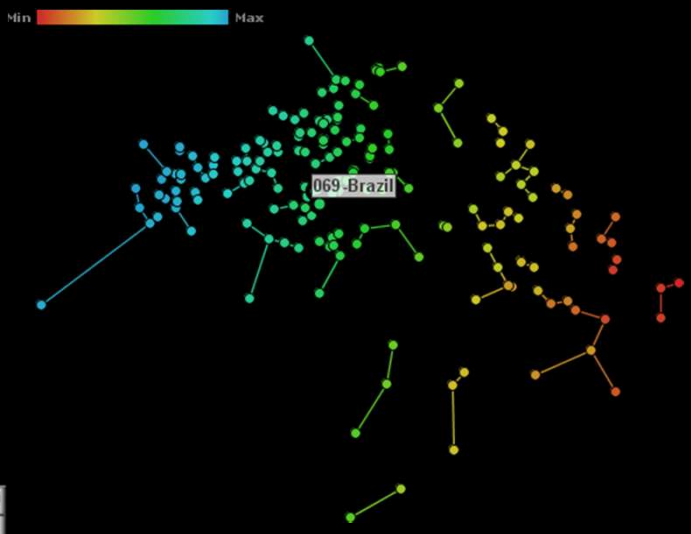
48

Exemplo de Projeção



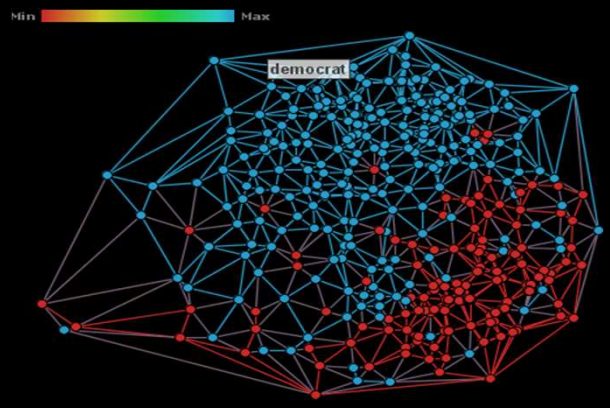
49

Exemplo de Projeção: IDH

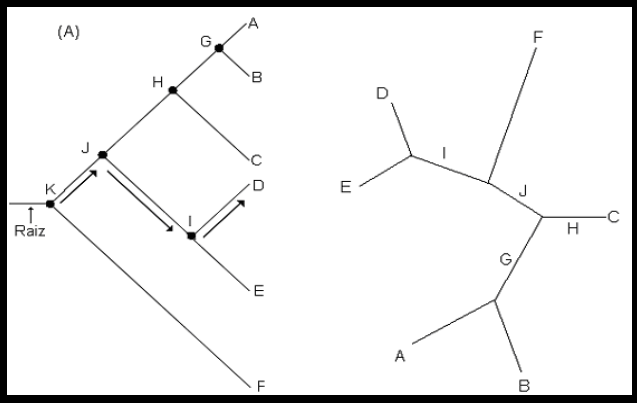


50

Exemplo de Projeção: Votação

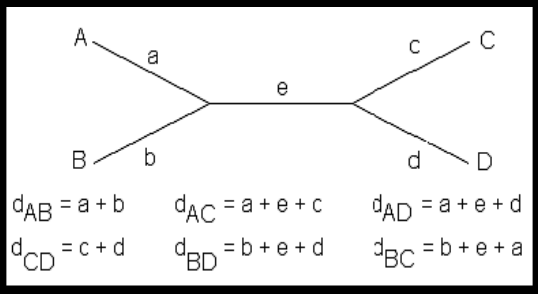


Point Placement by Phylogenetic Tree Construction Algorithms (N-J Trees)



Point Placement by Phylogenetic Tree Construction Algorithms (N-J Trees)

$$d_{AB} + d_{CD} \leq \max\left(d_{AC} + d_{BD}, d_{AD} + d_{BC}\right)$$



Algorithm Neighbor-joining

Input: distance matrix

1. Criate a star tree for n objects.
2. Iteration
   1. Select a node pair (i,j) with smaller Sij (branch size)

$$S_{ij} = \frac{1}{2(n-2)} \sum_{k=3}^{N} (D_{ik} + D_{jk}) + \frac{1}{2} D_{ij} + \frac{1}{n-2} \sum_{3 \leq m < n} D_{ij}$$

   2. Combine nodes i and j in a new node and calculate the branch size of the new node.

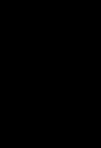
$$L_{ix} = \frac{D_{ij} + D_{iz} - D_{jz}}{2} \qquad L_{jx} = \frac{D_{ij} + D_{jz} - D_{iz}}{2}$$

Algorithm Neighbor-joining

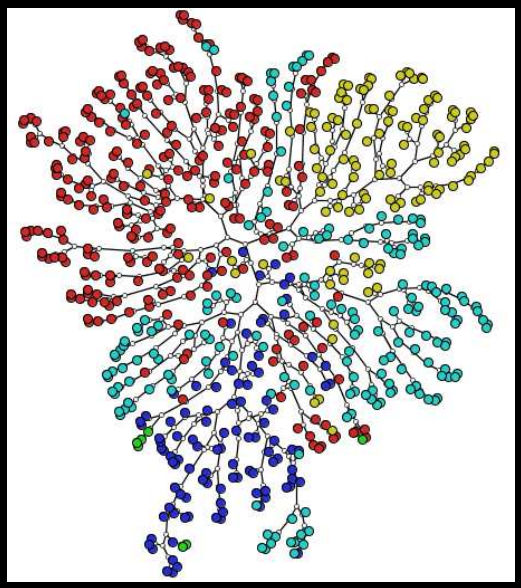
3. Calculate new distance matrix, computing the new distances from the new node to the remaining nodes.

$$D_{(i-j),k} = \frac{(D_{ik} + D_{jk})}{2} \qquad (3 \le k \le N)$$

4. Eliminate previous nodes i and j
5. If n>2 then iterate again.



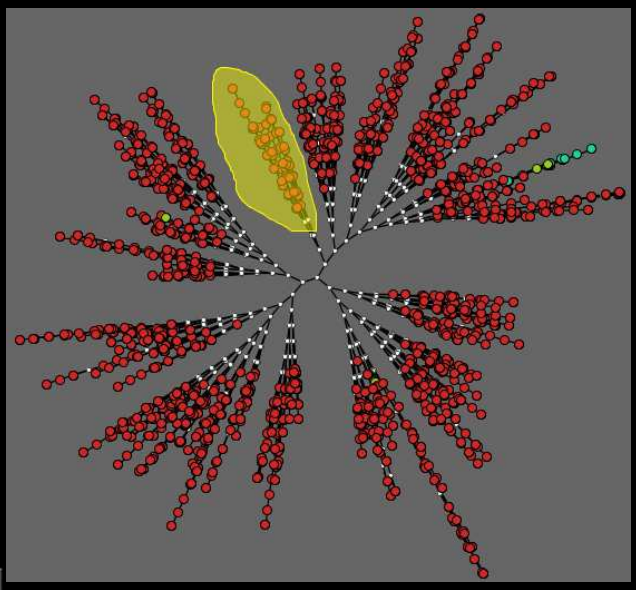
55



56

- Alternate view (N-J Tree)

57

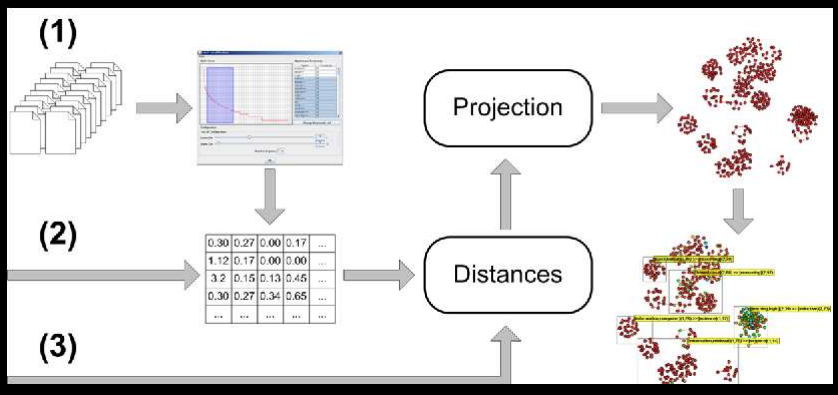


58

Context: Visual Data Mining

- Definition [Ankerst 2000]
  - step in process of knowledge discovery / extraction (KDD)
  - utilizes visualization as communication channel between computer and user
  - to support identification of new and interpretable patterns

Visualization by Projections



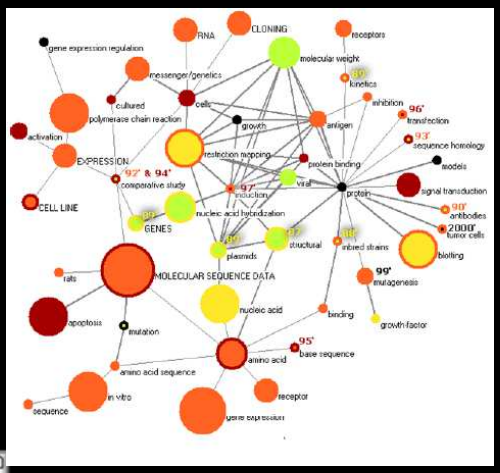
O caso de coleções de documentos

- Aplicações
  - Ensino/Pesquisa
  - Busca
  - Investigação

  - Patentes
  - Laudos médicos
  - Notícias

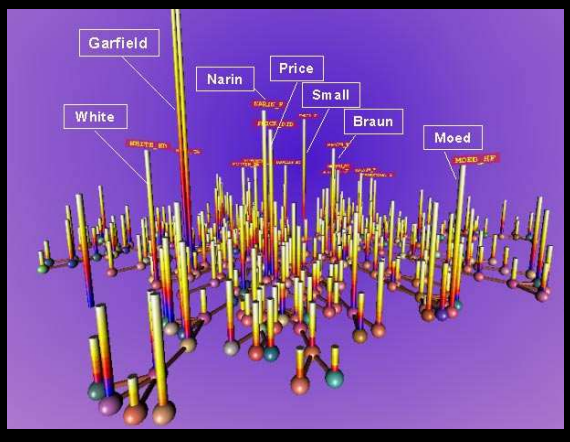
- Maps of text Collections
  - Based on Relationships (Borner & Chen)
    - Co-authorship, co-citation
  - Based on Content
    - Similarity and Grouping
    - Common underlying subject
    - ➔Topics

Relationships : Topic Busts and co-word



(Mane and Borner) 2004

Relationships : Citation and Co-citation



(Borner) (2003)

Content-based Text Mapping

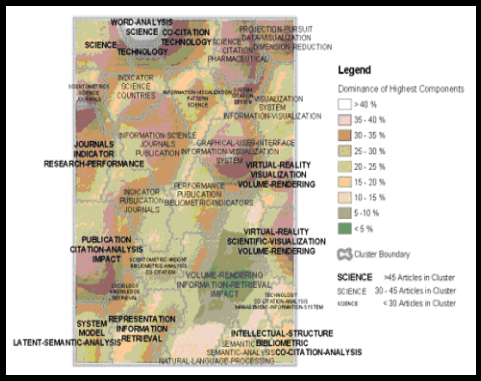
- Approach 1: Dimension reduction
  ex. MSD, SVD, PCA
- Approach 2: Point Placement (PP)
- Approach 3: Clustering
- Approach 4: Projections
  ex. FASPMAP, NNP, LSP

Content - based

(Skupin) (2002) (abstracts) SOM

SOM based

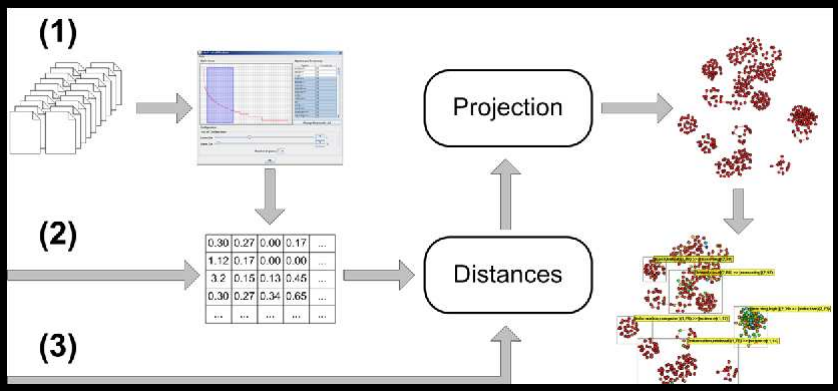
- **Self-Organization Maps (SOMs) cartográficos (ex. Skurpin 2002)**



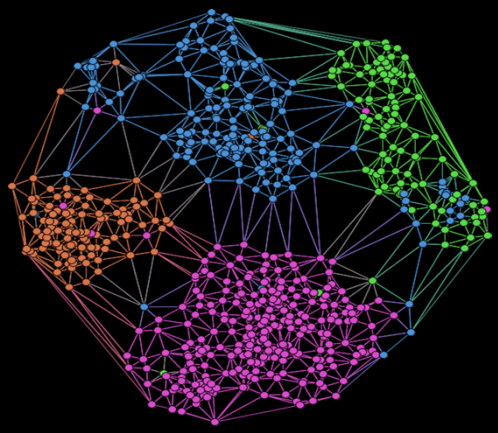
Content - based

(Dimensional Reduction)
News flash
IN-SPIRE
(PNL)

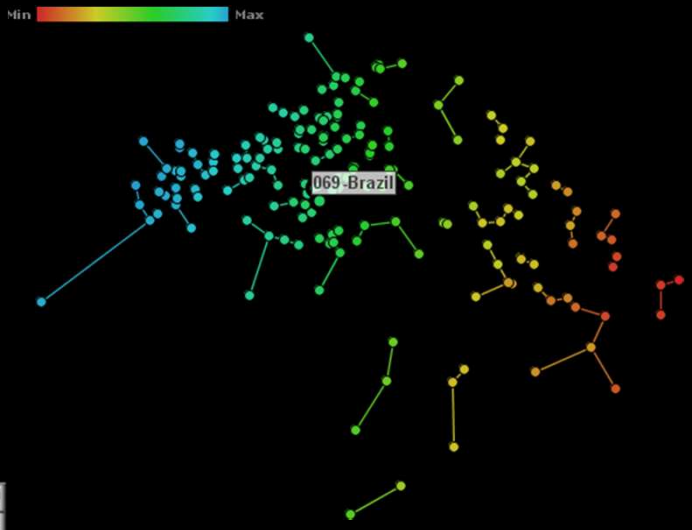
Content – based by Projections



Exemplo de Projeção

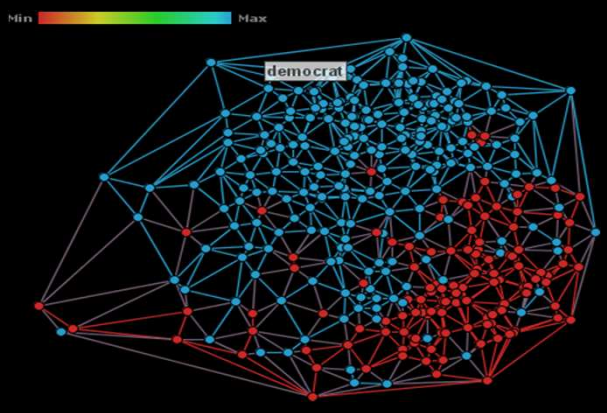


Exemplo de Projeção: IDH



71

Exemplo de Projeção: Votação



72

Continua na Parte II : variações e aplicações